

FAST CHANGE POINT ANALYSIS ON THE HURST INDEX OF PIECEWISE FRACTIONAL BROWNIAN MOTION

Pierre, R. BERTRAND^{1,2} *Pierre.Bertrand@math.univ-bpclermont.fr*

Mehdi FHIMA² *Mehdi.Fhima@math.univ-bpclermont.fr*

Arnaud GUILLIN² *Arnaud.Guillin@math.univ-bpclermont.fr*

¹ *INRIA Saclay*

² *Laboratoire de Mathématiques, UMR CNRS 6620*

É Université de Clermont-Ferrand II, France

Résumé: Dans cette présentation, nous introduisons une nouvelle méthode de détection de ruptures sur l'indice de Hurst, pour un mouvement brownien fractionnaire par morceaux. En premier lieu, nous définissons le modèle et le problème statistique. La méthode proposée est une transposition de la méthode FDpV à l'estimation de l'indice de Hurst. La méthode FDpV (dérivée filtrée avec *p-value*) a été introduite pour détecter des ruptures sur la moyenne par Bertrand *et al.* (2011). La statistique sous-jacente de la technologie FDpV est un nouvel estimateur de l'indice de Hurst, appelé statistique de Bernoulli des accroissements (IBS). À la fois les méthodes FDpV et IBS ont une complexité linéaire par rapport à la taille de la série d'observation, aussi bien en temps de calcul que pour la mémoire, donc également leur combinaison.

Mots clés: Détection de ruptures, Dérivée Filtrée, mouvement Brownien fractionnaire par morceaux, paramètre de Hurst, Statistique de Bernoulli des accroissements.

Abstract: In this presentation, we introduce a new method for change point analysis on the Hurst index for a piecewise fractional Brownian motion. We first set the model and the statistical problem. The proposed method is a transposition of the FDpV (Filtered Derivative with p-value) method introduced for the detection of change points on the mean in Bertrand *et al.* (2011) to the case of changes on the Hurst index. The underlying statistics of the FDpV technology is a new statistic estimator for Hurst index, so-called Increment Bernoulli Statistic (IBS). Both FDpV and IBS are methods with linear time and memory complexity, with respect to the size of the series. Thus the resulting method for change point analysis on Hurst index reaches also a linear complexity.

Keywords: Change point analysis, Filtered derivative with p-value method, Hurst parameter, Increment Bernoulli Statistic, piecewise fractional Brownian motion.

Introduction

Recent measurement methods allow us to record and to stock large data sets, so called "the data deluge". For instance, today technology allows recording of heartbeat series during 24 hours leading to data sets of size $n \geq 100,000$, and very high frequency (VHF) financial series leads to data size $n \geq 40,000$. Tomorrow, many other series will be recorded at VHF leading to millions of data.

Large or huge series with small meshes of time can be described as continuous time processes observed at discrete times. Such a stochastic process X belongs to a certain class of model, that is $X \in \mathcal{M} = \{X_\theta, \theta \in \Theta\}$, where Θ is a subset of \mathbb{R}^d and d is the dimension of the model. The

structural parameter θ is believed to provide relevant information on the system which generates the series, and statisticians have to estimate it.

A slightly different approach is based on change point analysis: The structural parameter θ is assumed to be piecewise constant with an unknown configuration of change τ . In this framework, the first task of statisticians is the estimation of the location of the change points and the second one could be the estimation of the structural parameters between change points. There is a huge literature on change point analysis and model selection since the fifties, see e.g. Basseville and Nikiforov (1993), Brodsky and Darkhovsky (1993), Csörgö and Horváth (1997), Birgé and Massart (2007) or Bertrand *et al.* (2011) and the references therein. However, most of the studies are devoted to change on the mean, on the variance or on the regression parameters. But relevant informations are also provided by the time dependence structure of the process, see e.g. Ayache and Bertrand (2011) and Khalfa *et al.* (2011). Fractional Brownian motion (fBm) is a paradigmatic example of such process, indeed fBm is a zero mean Gaussian process depending on two parameters: The Hurst index H linked to the time structure and a scale parameter σ .

In this presentation, we consider a simple model, that is a process X which is a piecewise fBm with an unknown configuration of changes. Moreover, we set us in the frame of huge datasets, and we focus our attention on time and memory complexity. These two reasons have lead us to propose a new change point procedure for detection of change on the Hurst index for a piecewise fBm. Our new procedure is the combination, on the one hand of the FDPV method, introduced in Bertrand *et al.* (2011) for fast and light detection of change on the mean, variance or regression parameter, and on the other hand of the Increment Bernoulli Statistic (IBS) a new estimator for Hurst index, which is a variation on the Increment Ratio Statistic (IRS) estimator introduced in Bardet and Surgailis (2010).

The rest of this paper is organized as follows: At first, in Section 1, we define our model of piecewise fBm. Next, in Section 2, we introduce a new fast and robust estimator of the Hurst index of fBm, namely the Increment Bernoulli Statistic. Then, in Section 3, we describe the transposition of the FDPV method to Hurst index.

1 Our model

We observe a process X at the discrete and regularly spaced time $t_i = i/n$, where $i = 0, \dots, n$. We assume the existence of a segmentation $\tau = (\tau_k)_{k=0, \dots, K+1}$, with $0 = \tau_0 < \tau_1 < \dots < \tau_K < \tau_{K+1} = n$, such that the restriction of the process X on each interval (τ_k, τ_{k+1}) for $k = 0, \dots, K$ is a fBm with Hurst index H_k and scale parameter σ_k . The integer K corresponds to the number of change points and $(K + 1)$ to the number of segments. Stress that K can be zero and in this case the process X is a fBm.

Let us precise that in our roadmap the process X should be almost surely with continuous paths. For this reason, the so-called piecewise fBm can not be defined by plugging a piecewise Hurst index into one of the representations of the fBm. Indeed, by doing so, the process X would almost surely have discontinuity at each change point on Hurst index, and method for detecting change on the mean will also be efficient for detecting change on Hurst index. Let us refer to Taqqu and Samorodnitsky (1994) as a reference book on the different representations of fBm and to Ayache and Taqqu (2005) for the construction of multi-fBm by plugging a continuous

time varying Hurst index into one of the fBm representations. A rather complicated solution to avoid the drawback of pathwise discontinuities due to Hurst index discontinuities, was proposed in Benassi *et al.* (2000) and cosigned by the first author. However, as point out by Antoine Ayache during private conversations held in 2004, for statistical applications, it suffices to cancel the discontinuity by adding a correction term. The same solution is also adopted in Bardet and Kammoun (2008) .

The model having being specified, we are concerned with change point analysis on the Hurst parameter, where the number of change K is unknown. There are few references on this problem. To our best knowledge, the only reference are Benassi *et al.* (2000) and Bardet and Kammoun (2008).

2 Increment Bernoulli Statistic for fBm

In this section, we investigate the properties of a new estimator of the Hurst parameter of fBm, namely the Increment Bernoulli Statistic (IBS). IBS is a variation on IRS which has been introduced by Surgailis *et al.* (2008) and applied to fBm by Bardet and Surgailis (2010). Both IRS and IBS are fast and robust estimator of the Hurst index. By fast we mean estimator with linear time complexity, and by robust we mean estimator with invariant scaling property. The choice of the IBS instead of the IRS is motivated by the fact that the IBS is bit less expensive, in terms of time complexity, than the IRS.

In the next section, the IBS is used as the underlying estimator for the FDpV method, see (1). For this reason, we define IBS for a every process X , even if we apply it to fBm in this section. Let X be a process observed at a family of discrete times t_k , we define the second order increments by

$$\Delta X(t_k) := X(t_{k+2}) - 2X(t_{k+1}) + X(t_k).$$

Then, the Increment Bernoulli Statistic (IBS) is based on the comparison of the signs of consecutive second order increments. The results of these comparisons will be equal to 1 if the consecutive second order increments have the same sign, and 0 otherwise. Hence, this explains the name of our new estimator, that is to say: Increment Bernoulli Statistic (IBS) which is given by

$$\text{IBS}_n(X) = \frac{1}{n-2} \sum_{k=0}^{n-3} \psi(\Delta X(t_k), \Delta X(t_{k+1}))$$

where $\psi(\cdot, \cdot)$ is described as follows $\psi(x, y) = 0$ if $\text{sign}(x) = \text{sign}(y)$ and 1 otherwise, where $\text{sign}(z)$ denotes the sign of z . Let us remark that IBS is scale invariant: Indeed, since $\psi(\sigma x, \sigma y) = \psi(x, y)$ for $\sigma > 0$, then $\text{IBS}_n(\sigma X) = \text{IBS}_n(X)$.

When X is a fBm, that is $X = B_H$ with a Hurst index $H \in (0, 1)$, then the IBS converges in distribution to a continuous monotonic increasing function $\Lambda(H)$ defined as follows

$$\begin{aligned} \Lambda(H) &:= \Lambda_0(\rho(H)) \\ \Lambda_0(r) &:= \frac{1}{\pi} \arccos(-r) \\ \rho(H) &= (-3^{2H} + 2^{2H+2} - 7) (8 - 2^{2H+1})^{-1} \end{aligned}$$

where $\rho(H) \in (-1, 1)$ represents the correlation between two successive second order increments. The graph of $\Lambda(H)$ is given in Figure 1. Then, due to the fact that $\Lambda(\cdot)$ is a reversible function, it is easy to deduce an estimator of the Hurst parameter H given by $\widehat{H}_n = \Lambda^{-1}(\text{IBS}_n(B_H))$. Furthermore, we note that the function $\phi(\cdot, \cdot) = \psi(\cdot, \cdot) - \Lambda(H)$ is a Hermite function with rank equal to 2. Then, by applying the Breuer-Major theorem, see *e.g.* Arcones (1994)[Theorem 4, p.2256] or Nourdin *et al.* (2010)[Theorem 1, p.2], we can deduce the following Central Limit Theorem (CLT):

$$\sqrt{n}(\text{IBS}_n(B_H) - \Lambda(H)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(H)),$$

where the sign $\xrightarrow{\mathcal{D}}$ means convergence in distribution and the asymptotic variance $\sigma^2(H)$ is given by

$$\sigma^2(H) = \sum_{j \in \mathbb{Z}} \text{cov}(\psi(\Delta B_H(t_0), \Delta B_H(t_1)), \psi(\Delta B_H(t_j), \Delta B_H(t_{j+1}))).$$

The main advantages of the IBS method are primarily its efficiency in terms of time and memory complexity, and secondarily its robustness with respect to scaling properties of the fBm. At first, we calculate by recurrence the second order increments $(\Delta B_H(t_k))_{0 \leq k \leq n-3}$. This first step is performed in time and memory complexity on $\mathcal{O}(n)$. Next, the computing of $\text{IBS}_n(B_H)$ requires roughly n tests + $n \times p_a(H)$ additions + 1 division, where $p_a(H) = \Lambda(H) \in (0, 1)$ corresponds to the probability that two consecutive second order increments have the same sign. Then, we apply the Newton algorithm to compute the inverse of the function $\Lambda(\cdot)$. Moreover, we note that the function $\psi(\cdot, \cdot)$ satisfy the scale invariant property, *i.e.* for all $C \in \mathbb{R}$, $\psi(C.X, C.Y) = \psi(X, Y)$. This means that the multiplication of B_H by any scaling coefficient C does not impact the estimation of the Hurst index, since $\text{IBS}_n(B_H) = \text{IBS}_n(C.B_H)$. Hence, this proves the robustness of the IBS method with respect to scaling.

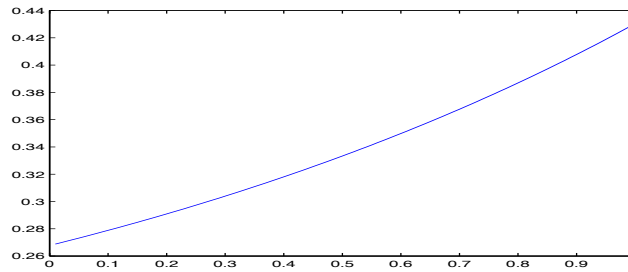


Figure 1: The graph of $\Lambda(H)$.

3 Filtered Derivative with p-Value method

In this section, we describe the Filtered Derivative with p-value method (FDp-V). First, we define the Filtered Derivative function. Next, we describe the two steps of the FDp-V method:

Step 1 is based on Filtered Derivative and select the potential change points, whereas Step 2 calculate the p-value associated to each potential change point, for disentangling right change points and false alarms.

At first, we note that $\Lambda(H)$ is a continuous monotonic increasing function of H , see Figure 1. So, the detecting of change points on the Hurst parameter H is equivalent to detecting change points on $\Lambda(H)$. Consequently, the estimator $\text{IBS}_n(B_H)$ of the parameter $\Lambda(H)$ is used as the underlying estimator for the FDpV method. We refer to Bertrand *et al.* (2011) and Bertrand and Fhima (2009) for the introduction of FDpV technology and its numerical efficiency. Let us stress that the choice of the direct estimator $\hat{H}_n = \Lambda^{-1}(\text{IBS}_n(B_H))$ as the underlying estimator for the FDpV would be more expensive in term of numerical complexity.

Filtered Derivative function

Let X be a piecewise fBm observed at a family of discrete times $t_j = j/n$, for $j = 0, \dots, n$. The Filtered Derivative for IBS is defined as the difference between the estimators of the parameter $\Lambda(H)$ computed on two sliding windows respectively at the right and at the left of the index k , both of size A , that is specified by the following function

$$D(k, A) = \text{IBS}(X, k, A) - \text{IBS}(X, k - A, A) \text{ for } k \in [A, n - A] \quad (1)$$

where

$$\text{IBS}(X, k, A) = A^{-1} \sum_{j=k+1}^{k+A} \psi(\Delta X(t_k), \Delta X(t_{k+1}))$$

is an estimator of $\Lambda(H)$ on the sliding box $[k + 1, k + A]$. It is easy to see that the Filtered Derivative function D is computed by recurrence with linear time and memory complexity. Eventually, this method consists on filtering data by computing the estimators of the parameter $\Lambda(H)$ before applying a discrete derivation. This construction explains the name given by Benveniste and Basseville (1984), the so-called Filtered Derivative method.

Step 1: Detection of potential change points

In order to detect the potential change points, we test the null hypothesis (\mathcal{H}_0) of no change in the Hurst parameter H against the alternative hypothesis (\mathcal{H}_1) indicating the existence of at least one change point

(\mathcal{H}_1) : There is an integer $K \in \mathbb{N}^*$ and $0 = \tau_0 < \tau_1 < \dots < \tau_K < \tau_{K+1} = n$ such that

$$H_1 = \dots = H_{\tau_1} \neq H_{\tau_1+1} = \dots = H_{\tau_2} \neq H_{\tau_2+1} = \dots = H_{\tau_K+1} = \dots = H_{\tau_{K+1}}.$$

where $H_j = H_{\tau_k}$ is the value of the Hurst parameter at $t_j \in [\tau_{k-1}/n, \tau_k/n)$.

Now, we fix a probability of type I error at level p_1^* , and we determine the corresponding critical value C_1 given by

$$\mathbb{P} \left(\max_{k \in [A: n-A]} |D(k, A)| > C_1 | \mathcal{H}_0 \text{ is true} \right) = p_1^*.$$

Of course, such a probability is usually not available, so that we only consider the asymptotic distribution of the maximum of $|D|$. Then, the change points $\tilde{\tau}_k$ is selected as a potential change

point if its local maxima satisfy $|D(\tilde{\tau}_k, A)| > C_1$. We remark through the graph of the function $|D|$ that there are not only the "right hats" (surrounded in green in Figure 2) which gives the right change points, but also false alarms (surrounded in black in Figure 2). Consequently, we have introduced another step in order to keep just the right change points.

Step 2: Elimination of false alarms

The list of potential change points $(\tilde{\tau}_1, \dots, \tilde{\tau}_{K_{\max}})$ obtained at step 1 contains right change points but also false alarms. In the second step, a test is carried out to remove the false alarms from the list of change points found at step 1. More precisely, for all potential change point $\tilde{\tau}_k$, we test whether the Hurst parameter is the same on the two successive intervals $(\tilde{\tau}_{k-1}/n, \tilde{\tau}_k/n)$ and $(\tilde{\tau}_k/n, \tilde{\tau}_{k+1}/n)$, or not. Formally, for all $1 \leq k \leq K_{\max}$, we apply the following hypothesis testing

$$(\mathcal{H}_{0,k}) : H_k = H_{k+1} \quad \text{versus} \quad (\mathcal{H}_{1,k}) : H_k \neq H_{k+1},$$

where H_k is the value of H on the segment $(\tilde{\tau}_{k-1}/n, \tilde{\tau}_k/n)$. By using this second test, we calculate new p-values $(\tilde{p}_1, \dots, \tilde{p}_{K_{\max}})$ associated respectively to each potential change points $(\tilde{\tau}_1, \dots, \tilde{\tau}_{K_{\max}})$. Then, we only keep the change points which have a p-value smaller than a critical level denoted p_2^* . By doing so, we obtain a subset $(\hat{\tau}_1, \dots, \hat{\tau}_{\hat{K}})$ of the first list which represents the estimators of the change points in the Hurst parameter of mBm.

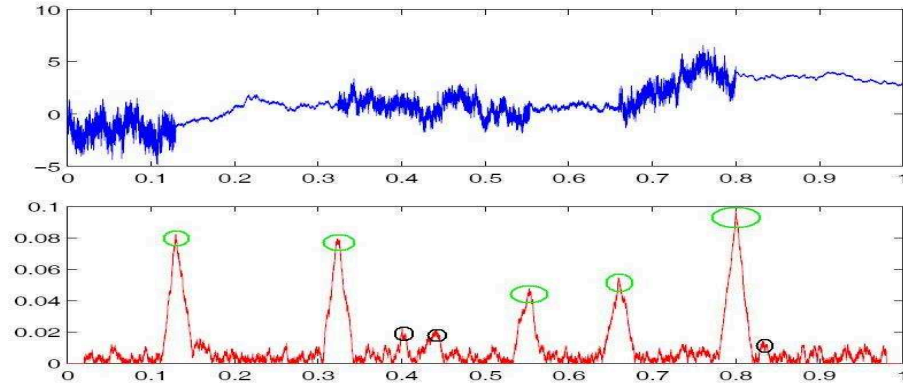


Figure 2: *Detection of potential change points. Above: Simulated piecewise fBm with five change points in the Hurst parameter. Below: Filtered Derivative function $|D|$.*

Conclusion

In conclusion, it appears that the combination of the FDpV and the IBS methods provides a fast (time) and cheap (memory) algorithm to the detection of change points on the Hurst parameter of piecewise fBm. So, this algorithm is adapted to segment random signals with large datasets. In future work, we will develop the FDpV + IBS method in order to detect abrupt changes on parameters of real data drawn from financial and physiological domains.

References

- [1] Antoch, J. and Hušková, M. (1994) Procedures for the detection of multiple changes in series of independent observations, *Asymptotic statistics (Prague, 1993)*, Contrib. Statist., 3–20. Physica, Heidelberg.
- [2] Arcones, M. A (1994) Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors, *Ann. Probab.*, 22, 2242–2274.
- [3] Ayache, A. and Bertrand, P.R. (2011) Discretization error of wavelet coefficient for fractal like process, to appear in *Advances in Pure and Applied Mathematics*.
- [4] Ayache, A. and Taqqu, M. S. (2005) Multifractional process with random exponent, *Publ. Mat.*, 49, 459–486.
- [5] Bardet, J. M. and Kammoun, I. (2008). Detecting abrupt changes of the long-range dependence or the self-similarity of a Gaussian process, *C. R. Math. Acad. Sci. Paris*, 346, 889–894.
- [6] Bardet, J. M. and Surgailis, D. (2010) Measuring Roughness of Random Paths by Increment Ratios, to appear in *Bernoulli*.
- [7] Basseville, M. and Nikiforov, I. V. (1993) *Detection of abrupt changes: theory and application*, Prentice Hall Inc, Englewood Cliffs, NJ.
- [8] Benassi, A., Bertrand, P. R., Cohen, S., Istas, J. (2000) Identification of the Hurst exponent of a Step Multifractional Brownian motion, *Statistical Inference for Stochastic Processes*, 3, 101–111.
- [9] Benveniste, A. and Basseville, M. (1984) Detection of abrupt changes in signals and dynamical systems: some statistical aspects, *Analysis and optimization of systems, Part 1 (Nice, 1984)*, volume 62 of *Lecture Notes in Control and Inform. Sci.*, 145–155, Springer, Berlin.
- [10] Bertrand, P. R. and Fhima, M. (2009) Filtered Derivative with p-Value Method for Multiple Change-Points Detection, *Proceeding of the 2nd International Workshop in Sequential Methodologies*.
- [11] Bertrand, P. R., Fhima, M. and Guillin, A. (2011) Off-line detection of multiple change points by the Filtered Derivative with p-Value method, to appear in *Sequential Analysis*.
- [12] Birgé, L. and Massart, P. (2007) Minimal penalties for Gaussian model selection, *Probab. Theory Related Fields*, 138(1-2), 33–73.
- [13] Brodsky, B. E. and Darkhovsky, B. S. (1993) *Nonparametric methods in change-point problems*, volume 243 of *Mathematics and its applications*, Kluwer Academic Publishers Group, Dordrecht.
- [14] Csörgö, M. and Horváth, L. (1997) *Limit Theorem in Change-Point Analysis*, J. Wiley, New York.
- [15] Khalfa, N., Bertrand, P. R., Boudet, G., Chamoux, A. and Billat, V. (2011), Heart Rate Regulation processed through wavelet analysis and change detection. Some case studies, *Submitted*.
- [16] Nourdin, I., Peccati, G. and Podolskij, M. (2010) Quantitative Breuer-Major Theorems, HAL : hal-00484096, version 2.
- [17] Samorodnitsky, G. and Taqqu, M. S. (1994) *Stable non-Gaussian random processes*, Chapman & Hall.
- [18] Surgailis, D., Teyssière, G. and Vaičiulis, M. (2008) The increment ratio statistic, *J. Multivariate Anal.*, 99, 510–541.